

IF 4071 Pembelajaran Mesin

Spesifikasi Tugas Besar 1

Decision Tree Learning

Tugas Besar 1a: Eksplorasi

- Buat Kelompok terdiri atas 3 orang
- Pelajari:
 - source code WEKA untuk ID3
 - source code WEKA untuk C4.5 (J48)
 - pelajari algoritma pada hal 56 buku Tom Mitchell.
- Carilah persamaan dan perbedaan ketiga algoritma berdasarkan source code tersebut dan jelaskanlah persamaan dan perbedaan tersebut dalam hal:
 - Penentuan atribut terbaik
 - Penanganan label dari cabang setiap nilai atribut
 - Bagaimana jika examples kosong di cabang tersebut
 - Bagaimana menangani atribut kontinu
 - Penanganan atribut dengan missing values
 - Pruning dan parameter confidence

Pengumpulan: Laporan (pdf), Minggu, 17 September 2017 Pukul 23.55 via Kuliah online K1 & K2 (kuliah.itb.ac.id)

Tugas Besar 1b: Create new classifier (1)

Bagian 1:

Menuliskan kode java untuk mengakses weka:

- Mulai dari load data (arrf dan csv)
- Remove atribut
- Filter : Resample
- Build classifier : DT
- Testing model given test set
- 10-fold cross validation, percentage split, training-test
- Save/Load Model
- Using model to classify one unseen data (input data)

Tugas Besar 1b: Create new classifier (2)

Bagian 2:

Membuat Classifier baru dengan menurunkan dari Classifier WEKA

Implementasi kelas baru pada weka:

1. myID3 (sesuai hal 56)
2. myC45 dengan mengakomodasi hal-hal berikut:
 - a) Overfitting training data: post pruning
 - b) Continuous-valued attribute: information gain dari candidate (maks 10)
 - c) Alternative measures for selecting attributes: gain ratio
 - d) Handling missing attribute value: most common target value

Testing data yang digunakan :

1. Data binary categorization weather (nominal, kontinu)
2. Data multiclass categorization iris

Pengumpulan: Minggu 1 Oktober 2017 via Kuliah online K1 & K2

LAPORAN (pdf) berisi:

- Source Code
- Hasil Eksekusi terhadap data tes
- Perbandingan dengan hasil ID3 & J48 weka
- Pembagian tugas setiap anggota kelompok